

(12) UK Patent Application (19) GB (11) 2 378 021 (13) A

(43) Date of A Publication 29.01.2003

(21) Application No 0209763.2

(22) Date of Filing 29.04.2002

(30) Priority Data

(31) 09853737 (32) 10.05.2001 (33) US

(71) Applicant(s)

Hewlett-Packard Company
(Incorporated in USA - Delaware)
3000 Hanover Street, Palo Alto,
California 94304, United States of America

(72) Inventor(s)

Pierre Perinet
Eric Peterson

(74) Agent and/or Address for Service

Carpmaels & Ransford
43 Bloomsbury Square, LONDON,
WC1A 2RA, United Kingdom

(51) INT CL⁷

G06F 11/34 17/30 17/40

(52) UK CL (Edition V)

G4A AFMG AFMT

(56) Documents Cited

EP 0494335 A2 WO 2002/029596 A1
WO 1998/038614 A2 US 5974457 A
US 5551025 A

(58) Field of Search

UK CL (Edition T) G4A AFMG AFMT
INT CL⁷ G06F 11/30 11/34 17/30 17/40
Other: Online: JAPIO, EPODOC, WPI

(54) Abstract Title

Archiving network usage data within a predetermined period

(57) A method for archiving data within a predetermined time interval, using an archiver connected to a data collector 18 having a list 26 of available data sets of an Internet Usage Manager 14, includes the steps of obtaining a list of available data sets from the data collector, reading data for a data set from the list, associating read data to the data set, converting all the read data to a predefined archive format once data from each available data set in the list is read, saving the converted data to a file 32, and waiting for a predetermined time to repeat the foregoing steps.

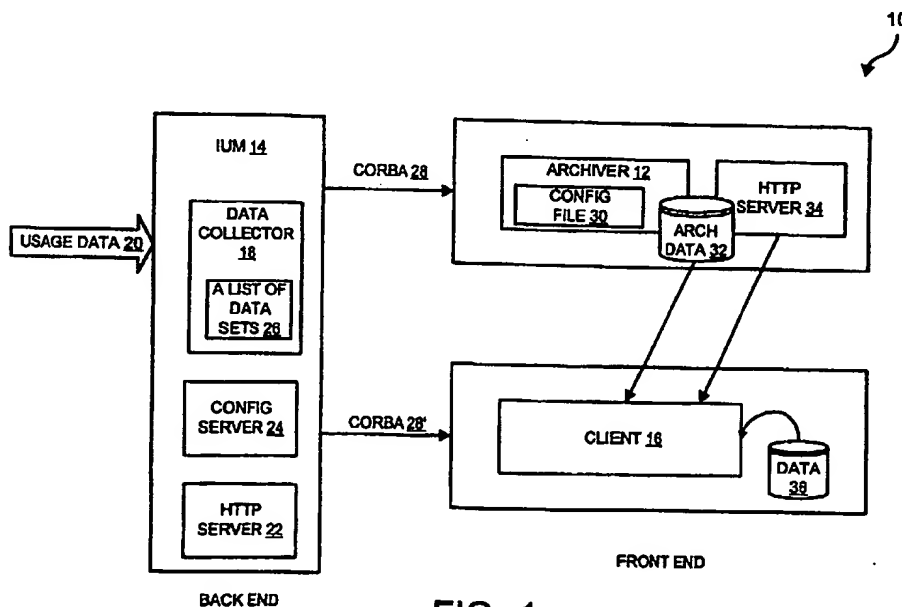


FIG. 1

GB 2 378 021 A

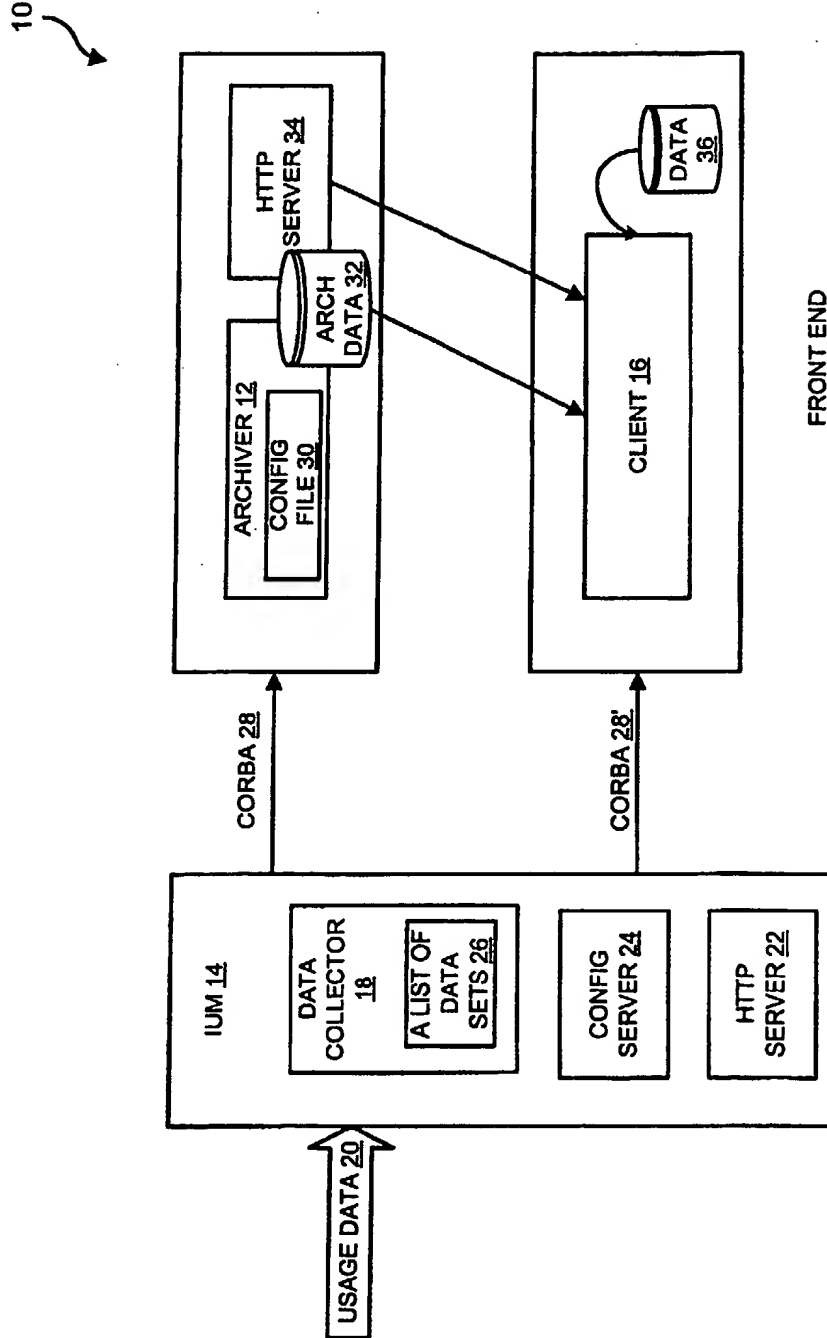


FIG. 1

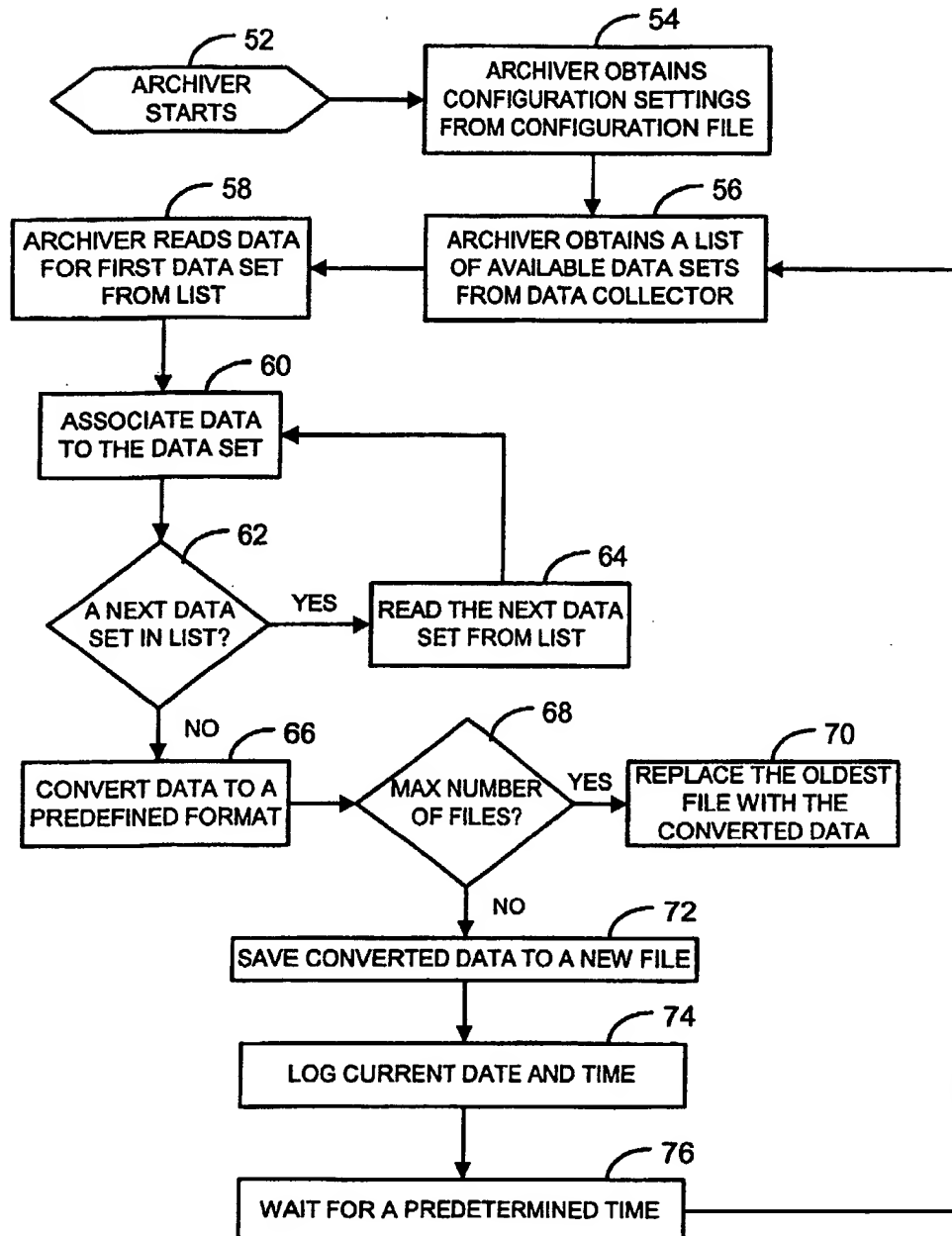


FIG. 2

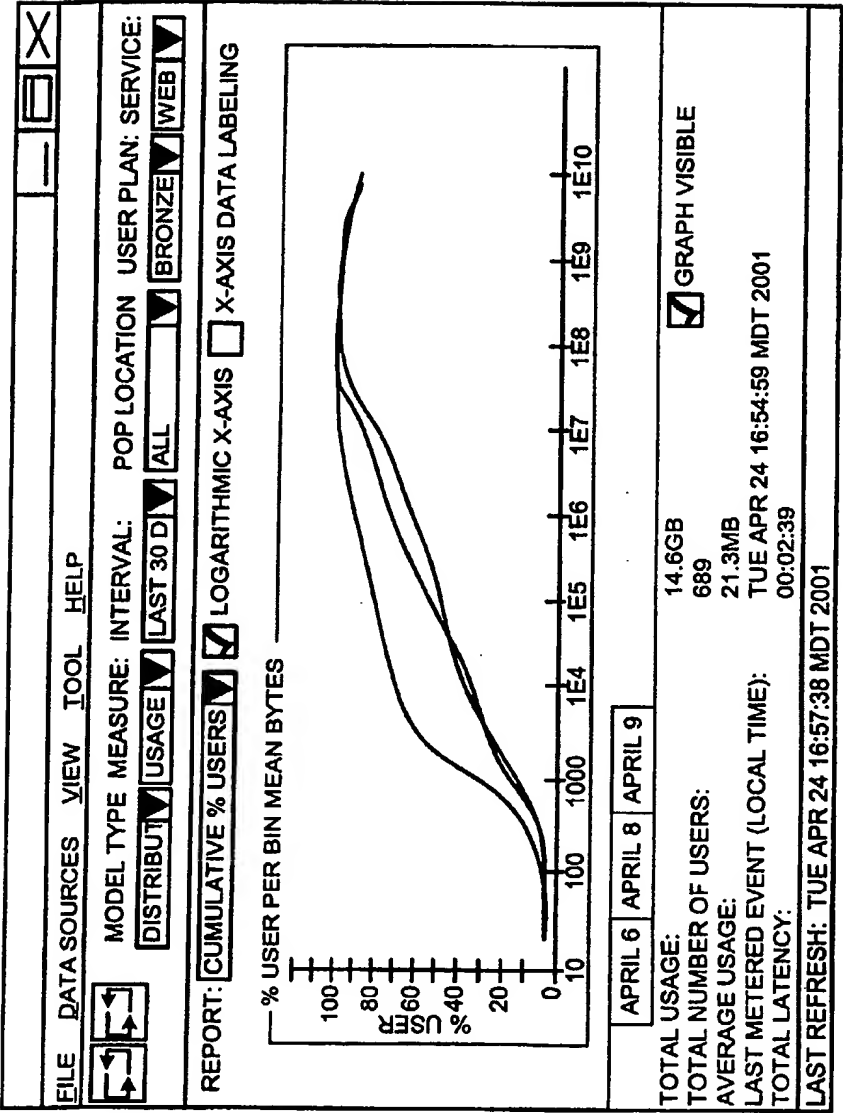


FIG. 3

1 Inventors: Pierre Perinet; Eric Peterson

2 METHOD AND SYSTEM FOR ARCHIVING DATA WITHIN A
3 PREDETERMINED TIME INTERVAL

4 The present invention generally relates to an improved method
5 and system for archiving data. More specifically, it relates to an improved
6 method and system for archiving data using an archiver connected to a data
7 collector having a list of available data sets of an Internet Usage Manager
8 within a predetermined time interval.

9 BACKGROUND OF THE INVENTIVE ART

10 Because Internet servers can provide valuable information about
11 their users, currently many software applications are designed to collect such
12 usage data. The data includes important information relating to items, such as
13 usage measure, geographical information, and user service requests. For
14 example, the data can provide valuable information for a business manager in

1 trying to understand the usage behavior of users, identify needs for new
2 services, managing the pricing of subscription plans, and determining profit
3 margin. All this information can provide managers with valuable marketing
4 tools.

5 Software applications for collecting usage data are generally
6 known as Internet Usage Managers ("IUM"). The present invention is
7 designed specifically for an IUM disclosed in a commonly owned U.S. Patent
8 application filed on April 12, 2000 entitled "Internet Usage Analysis System
9 And Method" bearing Serial No. 09/548,124 by Lee Rhodes,
10 assigned to HP company. This patent application is specifically incorporated
11 by reference herein.

12 An IUM typically includes a data collector for saving any user
13 data relating to the server. The data collector is set up to collect data
14 continuously. As a result, the usage data is not divided into various time
15 intervals, which could be extremely valuable information relating to the users.
16 For example, managers can track the behavior and traffic of the server in a
17 particular month, week or day. However, since the data collector collects
18 information continuously, this type of desired information is difficult to
19 achieve using the data collector of the IUM. As a result, there is need for an
20 improved IUM that collects data within a predetermined time interval selected
21 by users. More specifically, an archiver connected to the IUM that can
22 automatically archive data collected by the data collector according to a
23 predetermined time interval is needed.

24 BRIEF SUMMARY OF THE INVENTION

25 The present invention is directed to an improved method and
26 system for archiving data. More specifically, it relates to an improved method

1 and system for archiving data using an archiver connected to a data collector
2 having a list of available data sets of an Internet Usage Manager within a
3 predetermined time interval.

4 The present invention provides a method that includes the steps
5 of obtaining a list of available data sets from the data collector, reading data for
6 a data set from the list, associating read data to the data set, converting all the
7 read data to a predefined archive format once data from each available data set
8 in the list is read, saving the converted data to a file and waiting for a
9 predetermined time to repeat the foregoing steps.

10 The present invention also provides a system that includes an
11 Internet Usage Manager system for managing server usage data, a data
12 collector of the IUM for collecting the usage data, a list of data sets defined by
13 at least one category, an archiver for archiving the usage data within a
14 predetermined time interval, and a client for an user interface for selecting and
15 displaying the usage data.

16 DESCRIPTION OF THE DRAWINGS

17 FIGURE 1 is an architectural diagram of an implementation
18 using the present invention;

19 FIG. 2 is a flow chart illustrating the preferred functionality of a
20 method of the present invention;

21 FIG. 3 is an exemplary page displayed on the client using the data
22 archived within a predetermined time interval; and,

23 FIG. 4 is another exemplary page displayed on the client using
24 the data archived within a predetermined time interval.

GLOSSARY OF TERMS AND ACRONYMS

The following terms and acronyms are used throughout the detailed description:

Archiver. A computer for archiving data collected by the data collector of an Internet Usage Manager system within a predetermined time interval.

Archive. A single file containing one or more separate files plus information into a format, such as XML or Binary, that allows them to be extracted by a suitable program.

Binary data. A file format for digital data encoded as a sequence of bits but not consisting of a sequence of printable characters (text). The term is often used for executable machine code.

Common Object Request Broker Architecture ("CORBA"). An Object Management Group ("OMG") specification which provides the standard interface definition between OMG-compliant objects.

Data Collector. A module in the Internet Manger Usage system that continuously collect usage data of the server.

Extensible Markup Language ("XML"). An initiative from the W3C defining an "extremely simple" dialect of SGML suitable for use on the World-Wide Web.

Hyperlink. A navigational link from one document to another, from one portion (or component) of a document to another, or to a Web resource, such as a Java applet. Typically, a hyperlink is displayed as a highlighted word or phrase that can be selected by clicking on it using a mouse to jump to the associated document or document portion or to retrieve a particular resource.

1 HTML (HyperText Markup Language). A standard coding
2 convention and set of codes for attaching presentation and linking attributes to
3 informational content within documents. (HTML 2.0 is currently the primary
4 standard used for generating Web documents.) During a document authoring
5 stage, the HTML codes (referred to as "tags") are embedded within the
6 informational content of the document. When the Web document (or HTML
7 document) is subsequently transferred from a Web server to a browser, the
8 codes are interpreted by the browser and used to display the document.
9 Additionally in specifying how the Web browser is to display the document,
10 HTML tags can be used to create links to other Web documents (commonly
11 referred to as "hyperlinks"). For more information on HTML, see Ian S.
12 Graham, The HTML Source Book, John Wiley and Sons, Inc., 1995 (ISBN
13 0471-11894-4).

14 Hyper Text Transport Protocol ("HTTP"). The standard World
15 Wide Web client-server protocol used for the exchange of information (such as
16 HTML documents, and client requests for such documents) between a browser
17 and a Web server. HTTP includes a number of different types of requests,
18 which can be sent from the client to the server to request different types of
19 server actions. For example, a "GET" request, which has the format GET
20 <URL>, causes the server to return the document or file located at the specified
21 URL.

22 Internet. A collection of interconnected or disconnected networks
23 (public and/or private) that are linked together by a set of standard protocols
24 (such as TCP/IP and HTTP) to form a global, distributed network. (While this
25 term is intended to refer to what is now commonly known as the Internet, it is
26 also intended to encompass variations which may be made in the future,
27 including changes and additions to existing standard protocols).

1 Internet Usage Manager ("IUM"). A computer implemented
2 system for managing usage data of the server.

3 Object Management Group ("OMG"). A consortium aimed at
4 setting standards in object-oriented programming.

5 Object-Oriented Programming. The use of a class of
6 programming languages and techniques based on the concept of an "object"
7 which is a data structure (abstract data type) encapsulated with a set of routines
8 that operates on the data. Operations on the data can only be performed via the
9 routine sets. These routine sets are common to all objects that are instances of
10 a particular "class". As a result, the interface to objects is well defined, and
11 allows the code implementing the routine sets to be changed so long as the
12 interface remains the same.

13 Standard Generalized Markup Language ("SGML"). A generic
14 markup language for representing documents. SGML is an International
15 Standard that describes the relationship between a document's content and its
16 structure. SGML allows document-based information to be shared and re-used
17 across applications and computer platforms in an open, vendor-neutral format.

18 URL (Uniform Resource Locator). A unique address which fully
19 specifies the location of a file or other resource on the Internet or a network.
20 The general format of a URL is protocol://machine address:port/path/filename.

21 Usage Data. Data collected by the IUM relating, among other
22 things, to information on users, sessions and usage.

23 World Wide Web ("Web"). Used herein to refer generally to both
24 (i) a distributed collection of interlinked, user-viewable hypertext documents
25 (commonly referred to as Web documents or Web pages) that are accessible via
26 the Internet, and (ii) the client and server software components which provide
27 user access to such documents using standardized Internet protocols. Currently,

1 the primary standard protocol for allowing applications to locate and acquire
2 Web documents is HTTP, and the Web pages are encoded using HTML.
3 However, the terms "Web" and "World Wide Web" are intended to encompass
4 future markup languages and transport protocols which may be used in place of
5 (or in addition to) HTML and HTTP.

6 DETAILED DESCRIPTION

7 Broadly stated, the present invention is directed to an improved
8 method and system for archiving data. The method and system provide an
9 archiver for archiving data collected by a data collector of an IUM within a
10 predetermined time interval. Because the archiver can be set up separately to
11 collect data within a predetermined time interval, usage data can be displayed
12 to users according to years, months, weeks, days, hours, minutes and seconds.
13 Usage information provided in such a context can greatly assist managers in
14 understanding the behavior and usage of users.

15 An architectural diagram of an implementation using the present
16 invention with an IUM is shown in FIG. 1, and indicated generally at 10. An
17 archiver 12 is connected between an IUM 14 and a client 16. The IUM 14 is a
18 computer for managing server statistical usage data, and includes a data
19 collector 18 for collecting usage data 20 of users using a HTTP server 22 with
20 specific server configurations 24. It should be noted that the preferred
21 implementation of the present invention is for use with Internet servers (e.g.
22 HTTP servers). However, other servers such as intranet or network servers can
23 be used. These other implementations with the use of other types of servers are
24 within the scope of the present invention. A list of a plurality of data sets 26
25 defined by at least one category is also included with the IUM. Data collected
26 by the data collector is grouped according to each data set in the lists.

1 The IUM 14 is preferably linked to the archiver 12 and the client
2 16 via a CORBA connection 28, 28'. Using the settings defined in the
3 configuration file 30, the archiver 12 automatically archives data from the data
4 collector within a predetermined time interval. The archived data 32 is then
5 saved locally. Because the archiver 12 also services the client 16, a HTTP
6 server 34 is preferably used for storing the archived data 32 and servicing the
7 client.

8 The client 16, on the other hand, is a user interface for displaying
9 the usage data collected by the archiver 12 or the data collector 18 of the IUM
10 14. If a user desires usage data within a predetermined time interval, the client
11 gathers the needed data from the archive. However, the client 16 can also
12 access the data collector 18. In fact, the client can access data 36 saved locally
13 on the client.

14 Although it is shown that the archiver 12, the IUM 14 and the
15 client 16 are located on different computers, they can be combined together in
16 any number of computers depending on the preferred implementation. In fact,
17 as known by those of ordinary skill, the network topology of the present
18 invention can be implemented in various ways, and these other
19 implementations are within the scope of the present invention.

20 Turning to an important aspect of the preferred embodiment of
21 the present invention, a flow chart of the preferred functionality of a
22 configuration method is shown in FIG. 2, and indicated generally at 50. At the
23 start of the archiver (Block 52), it first obtains configuration settings from the
24 configuration file (Block 54), which preferably includes a URL of the data
25 collector, a maximum number of files allowed and a predetermined time
26 interval for collecting the data. Other information and settings relating to the

1 archiver can also be included in the configuration file. Next, the archiver
2 obtains a list of available data sets from the data collector (Block 56).

3 The archiver will archive data associated with each data set in the
4 list. More specifically, the archiver reads data for the first data set from the list
5 (block 58). The data will be associated with this first data set (block 60). It is
6 next determined whether there is a next data set in the list (block 62). If there
7 is, the archiver will continue to read the next data set from the list (block 64)
8 and an association of the data will similarly be made with this next data set
9 (Block 60). The archiver continues reading the data for each data set from the
10 list (block 64) until a next data set is no longer available in the list (block 62).
11 Once all the data has been read for every data set in the list (block 62), the read
12 data is converted to a predefined archive format (block 66). The preferred
13 archive format is preferably either binary or XML. However, any archive
14 format can be used, and they are within the scope of the present invention.

15 Before the converted data is saved locally on the archiver, it is
16 first determined whether there is already the maximum number of files allowed
17 (Block 68). A maximum number of files allowed is preferably provided for in
18 the configuration file to avoid having endless megabytes of files that
19 unnecessarily take up memory of the archiver. Although it is not essential to
20 the method, it is preferred that some type of maintenance for the archiver is
21 included with the present method. If the maximum number of files allowed are
22 already in the system (block 68), the converted data will be saved by replacing
23 it in the oldest file allowed in the system (block 70). On the other hand, if the
24 maximum number of files allowed are not already in the system (block 68), the
25 converted data will then be saved onto a new file (block 72). Once the
26 converted data has been saved, the archiver will log the date and time to
27 indicate when the data was collected and archived from the data collector of the

1 IUM (block 74). The archiever will then wait for a predetermined time from
2 the last logged date and time (block 76) to restart the process from the step of
3 obtaining a list of available data sets from the data collector (block 56).

4 Exemplary pages displayed on the client using the data archived
5 within a predetermined time interval by the archiver is shown in FIGS. 3 and 4.
6 Because the archiver collects data according to a predetermined time interval,
7 such as once a day, various histograms, graphs and charts can be viewed within
8 a time interval set by users. Near the top of the screen, users can choose
9 specific parameters relating to categories of model type, measure, time interval,
10 geographical location, user plan or user service. In this example, the data set is
11 defined as "distribution" for model type, "usage" for measure, "last 30 days" for
12 time interval, "all" for geographical location, "bronze" for user plan and "web"
13 for user service.

14 A single data set with these specific parameters is defined in the
15 list. For example, if the category time interval is changed to last week with the
16 other categories remaining the same, another data set is defined for these
17 parameters. Thus, in the list of the IUM, the preferred embodiment has
18 thousands of data sets. However, as shown, both examples in FIGS. 3 and 4
19 define the same data set, and the only difference is the type of report the user
20 selected. More specifically, a cumulative % users for FIG. 3 and a historical
21 summary for FIG. 4.

22 Because the archiver automatically archives the data within a
23 predetermined time interval, the client is able to view data according to various
24 time interval, which is not available through the data collector of the IUM. As
25 a result, managers can view usage data that are more relevant to marketing.
26 The archiver provides an additional dimension to IUM systems for providing
27 an improved and valuable marketing tool.

1 From the foregoing description, it should be understood that an
2 improved method and system for archiving data for use with an IUM has been
3 shown and described, which has many desirable attributes and advantages. The
4 method and system provide users with data according to a predetermined time
5 interval on the IUM. As a result, a time interval category is added to the IUM
6 systems, which provides important data information. Thus, the archiver
7 provides an additional dimension for data archiving in IUM systems.

8 While various embodiments of the present invention have been
9 shown and described, it should be understood that other modifications,
10 substitutions and alternatives are apparent to one of ordinary skill in the art.
11 Such modifications, substitutions and alternatives can be made without
12 departing from the spirit and scope of the invention, which should be
13 determined from the appended claims.

14 Various features of the invention are set forth in the appended
15 claims.

WHAT IS CLAIMED IS:

- 1 1. A method for archiving data using an archiver connected
2 to a data collector having a list of available data sets of an Internet Usage
3 Manager within a predetermined time interval, comprising the steps of:
4 obtaining a list of available data sets from the data collector;
5 reading data for a data set from the list;
6 associating read data to the data set;
7 converting all the read data to a predefined archive format once
8 data from each available data set in the list is read;
9 saving the converted data to a file; and,
10 waiting for a predetermined time to repeat the foregoing steps.
- 1 2. The method according to claim 1 wherein said archiver is
2 configured for use with an Internet Usage Manager system.
- 1 3. The method according to claim 1 wherein said data set
2 includes any one from the category of model type, measure, time interval,
3 geographical location, user plan or user service.
- 1 4. The method according to claim 1 wherein the converted
2 data is in Binary Language or Extensible Markup Language.
- 3 5. The method according to claim 1 wherein prior to the step
4 of obtaining a list of available data sets further comprises the step of obtaining
5 configuration settings from a configuration file for the archiver.
- 1 6. The method according to claim 5 wherein said
2 configuration file comprises an uniform resource locator of the data collector, a
3 maximum number of files allowed, and a predetermined time interval for
4 collecting the data.

1 7. The method according to claim 1 wherein said step of
2 reading data for a data set from the list further comprises the step of:
3 reading a first data set from the list;
4 associating read data to the first data set;
5 determining whether there is a next data set available in the list;
6 reading the next data set from the list when there is a next data set
7 available in the list; and,
8 converting the read data once there is no next data set available in
9 the list.

1 8. The method according to claim 1 wherein said step of
2 saving the converting data to a file further comprises the steps of:
3 determining whether there is already a maximum number of files
4 allowed;
5 replacing the oldest file with the converted data if there is already
6 a maximum number of files allowed;
7 saving the converted data to a new file if there is not a maximum
8 number of files allowed.

1 9. The method according to claim 1 wherein prior to said
2 step of waiting for a predetermined time further comprising the step of logging
3 the current date and time.

1 10. A computer program product comprising a computer
2 usable medium having computer readable program codes embodied in the
3 medium that when executed causes a computer to:
4 obtain a list of available data sets from the data collector;
5 read data for a data set from the list;
6 associate read data to the data set;

7 convert all the read data to a predefined archive format once data
 8 for each available data set in the list is read;
 9 save the converting data to a file; and,
 10 wait for a predetermined time to repeat the process.

1 11. An archiver for archiving data within a predetermined
 2 time interval connected to an Internet Usage Manager system having a data
 3 collector for collecting available data for a list of data sets to:

4 obtain a list of available data sets from the data collector;
 5 read data for a data set from the list;
 6 associate read data to the data set;
 7 convert all the read data to a predefined archive format once data
 8 for each available data set in the list is read;
 9 save the converting data to a file; and,
 10 wait for a predetermined time to repeat the process.

1 12. A system for archiving data using an archiver connected to
 2 a data collector having a list of available data sets of an Internet Usage
 3 Manager within a predetermined time interval, comprising:

4 an Internet Usage Manager system for managing usage data of a
 5 server;
 6 a data collector of said Internet Usage Manager system for
 7 collecting the usage data;
 8 a list of data sets defined by at least one category;
 9 an archiver for archiving the usage data within a predetermined
 10 time interval;
 11 a client for an user interface for selecting and displaying the
 12 usage data.



INVESTOR IN PEOPLE

Application No: GB 0209763.2
Claims searched: 1-12

Examiner: Geoff Western
Date of search: 20 November 2002

Patents Act 1977 Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.T): G4A (AFMG, AFMT)

Int Cl (Ed.7): G06F 11/30, 11/34, 17/30, 17/40

Other: Online: JAPIO, EPODOC, WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
X	EP 0494335 A2 (HEWLETT PACKARD) N.b. pages 3-5, Fig 8	1,2,10-12,at least
X,P	WO 2002/029596 A1 (CERBERIAN et al) N.b. page 8, Fig2, module 216	1,2,10-12,at least
X	WO 1998/038614 A2 (WEBTRENDS) N.b. pages 5-14	1,2,10-12,at least
X	US 5974457 A (DAUGHERTY et al) N.b. pages 3-5	1-3,10-12,at least
X	US 5551025 A (BRAZIER et al) N.b. pages 2-4	1,2,10-12,at least

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.